【云南财经大学博士论丛】

Bayesian Criterion-Based Model Selection in Structural Equation Models

结构方程模型的选择方法研究

——一种基于贝叶斯准则的模型选择方法

Li Yunxian

西南交通大学出版社
Http://press.swjtu.edu.cn

云南财经大学博士论丛

Bayesian Criterion-Based Model Selection in Structural Equation Models

结构方程模型的选择方法研究
——一种基于贝叶斯准则的模型选择方法

Li Yunxian

西南交通大学出版社
·成 都·

图书在版编目（CIP）数据

结构方程模型的选择方法研究：一种基于贝叶斯准则的模型选择方法 / 李云仙著. —成都：西南交通大学出版社，2011.12
（云南财经大学博士论丛）
ISBN 978-7-5643-1487-3

Ⅰ. ①结… Ⅱ. ①李… Ⅲ. ①社会科学－统计模型－线性模型－选择方法－研究 Ⅳ. ①C32

中国版本图书馆 CIP 数据核字（2011）第 236448 号

云南财经大学博士论丛
结构方程模型的选择方法研究
——一种基于贝叶斯准则的模型选择方法
李云仙

责任编辑	刘　立
特邀编辑	罗　旭
封面设计	何东琳设计工作室
出版发行	西南交通大学出版社 （成都二环路北一段 111 号）
发行部电话	028-87600564　028-87600533
邮政编码	610031
网　　址	http: //press.swjtu.edu.cn
印　　刷	四川经纬印务有限公司
成品尺寸	146 mm×208 mm
印　　张	4.75
字　　数	172 千字
版　　次	2011 年 12 月第 1 版
印　　次	2011 年 12 月第 1 次
书　　号	ISBN 978-7-5643-1487-3
定　　价	20.00 元

图书如有印装质量问题　本社负责退换

Preface

Structural equation models (SEM) are widely used in behavioral, educational, medical, and social sciences. In the past few years, the growth of SEM is very rapid. New models and statistical methods have been developed for better analysis of more complex data structures in substantive research. These include but are not limited to: nonlinear structural equation models; structural equation models with mixed continuous and ordered categorical data; multi-level structural equation models and finite mixture structural equation models.

Also many methods have been developed to analyze SEMs. One popular method is the Bayesian approach. An important issue in the Bayesian analysis of SEMs is model selection. In the literature, Bayes factor and deviance information criterion (DIC) are commonly used statistics for Bayesian model selection. However, as commented in Chen et al(2004), Bayes factor relies on posterior model probabilities, in which proper prior distributions are needed. And specifying prior distributions for all models under consideration is usually a challenging task, in particular when the model space is large. In addition, it is well known that Bayes factor and posterior model probability are generally sensitive to the choice of the prior distributions of the parameters. Furthermore the computational burden of Bayes factor is heavy. Alternatively, criterion-based methods are attractive in the sense that they do not require proper prior distributions in general, and the computation is quite simple. One of commonly used criterion-based methods is DIC, which however assumes the posterior mean to be a good estimator. For some models like the mixture SEMs,

WinBUGS does not provide the DIC values. Moreover, if the difference in DIC values is small, only reporting the model with the smallest DIC value may be misleading.

Motivated by the limitations of the Bayes factor and DIC above, this book aims to give a complete and self-contained presentation of a Bayesian criterion-based model selection method, called the L_v measure, for structural equation models. It is a combination of the posterior predictive variance and bias, and can be viewed as a Bayesian goodness-of-fit statistic. The calibration distribution of the L_v measure, defined as the prior predictive distribution of the difference between the L_v measures of the candidate model and the criterion minimizing model, is discussed to help understanding the L_v measure in detail. The computation of the L_v measure is quite simple, and the performance is satisfactory. Thus, it is an attractive model selection statistic.

In this book, the application of the L_v measure to various kinds of SEMs will be studied, and some illustrative examples will be conducted to evaluate the performance of the L_v measure for model selection of SEMs. To compare different model selection methods, Bayes factor and DIC will also be computed. Moreover, different prior inputs and sample sizes are considered to check the impact of the prior information and sample size on the performance of the L_v measure.

This book addresses a very large public as it includes graduate students and academic researchers in statistics and applied statistics, in behavioral, educational, medical, and social sciences. Readers who have mastered the material in this book will see how the Bayesian criterion-based method can be extended in a structural equation models to provide satisfactory result, more general and able to solve problems in a more complicated models. They will indeed have a new approach giving a more competitive knowledge related to the complexity of real-life problems.

Contents

Chapter 1 Introduction

1.1 Overview

Structural equation models (SEMs) are very popular in analyzing relationships among observed and latent variables. Nowadays, SEMs have been applied to many fields, including but not limited to business, marketing, education, medicine, psychology and sociology. One of the main objectives of these applications is to search a good SEM that can reveal the relationships among covariates, observed and latent variables. Hence, model selection is an important issue in analyzing SEMs. Moreover, as explained in Lee (2007, Chap. 5), hypothesis testing can be treated as a model selection problem.

Recently, the Bayesian approach for analyzing SEMs has received much attention, see Schines et al. (1999), Dunson (2000), Ansari et al. (2002), Lee (2007), Lee et al. (2010), Cai et al. (2010), and references therein. So far, the most widely used Bayesian model selection statistics are Bayes factor (Kass and Raftery, 1995) and the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). It is well known that for complex statistical models, the computation of Bayes factor is difficult (DiCiccio et al., 1997). Gelman and Meng (1998) developed an efficient algorithm, namely the path sampling, to compute the normalizing constant of a probability density function. This algorithm has been applied to compute the Bayes factors of many complex SEMs (see for example, Lee and Song, 2002, 2003a; Song and Lee, 2007, 2008, and the references

therein). Like the Bayesian Information Criterion (BIC), DIC takes into account the number of unknown parameters in the model. As the software WinBUGS (Spiegelhalter et al., 2003) provides the DIC values for most SEMs, the application of DIC is convenient.

While Bayes factor and DIC have some nice features, they have limitations. It is well-known that Bayes factor requires proper prior distributions of the parameters. In fact, it will favor the competitive model M_0 if the prior of the parameters in model M_1 has a very large spread so as make it non-informative. This is known as the "Bartletts Paradox". Moreover, for competitive models M_0 and M_1, such as multilevel SEMs with very different structures, it is difficult to find a direct path to link them when applying the path sampling. Under these cases, some auxiliary models may have to be used in computing the Bayes factor (see Lee, 2007). This will increase the computational burden. For DIC, it assumes the posterior mean to be a good estimator; and for some models (for example, the mixture SEMs), WinBUGS does not give the DIC values. Moreover, if the difference in DIC values is small, only reporting the model with the smallest DIC value may be misleading. In this book, motivated by the above limitations of the Bayes factor and DIC, we propose an attractive Bayesian statistic for model selection for different kinds of SEMs.

The proposed Bayesian statistic, called the L_v measure, is a criterion-based method that does not require proper prior distributions of the parameters. It will be shown that the computational burden involved is light, and the statistic can be obtained conveniently via observations simulated for the Bayesian estimation. Basically, the L_v measure involves two components. The first component is related to the reliability of the prediction, and the second component measures the discrepancy between the prediction and the observed data. Hence, it can be used to examine the goodness-of-fit of the model to the observed data. We will also consider the

calibration distribution of the L_v measure, which will allow us to compare two competing models in more details.

1.2 L_v Measure for Model Selection

To define the criterion, some notations will be used. Let $\boldsymbol{Y}^{\text{obs}} = (\boldsymbol{y}_1^{\text{obs}}, \cdots, \boldsymbol{y}_n^{\text{obs}})$ be a matrix of observations, and $\boldsymbol{Y}^{\text{rep}} = (\boldsymbol{y}_1^{\text{rep}}, \cdots, \boldsymbol{y}_n^{\text{rep}})$ be a matrix of replications, which has the same distribution with $\boldsymbol{Y}^{\text{obs}}$, and $\boldsymbol{\theta}$ be a vector that contains all the unknown parameters in the given model.

Gelfand and Ghosh (1998) proposed a minimum posterior predictive loss approach for model selection. They obtained the criterion by minimizing posterior loss for a given model and then, for models under consideration, selecting the one which minimizes this criterion. To define the loss, let $\boldsymbol{a} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ be the action matrix which is an estimate trying to accommodate both $\boldsymbol{Y}^{\text{obs}}$, and what we predict for $\boldsymbol{Y}^{\text{rep}}$. For the ith observation in $\boldsymbol{Y}^{\text{obs}}$, let $L(\boldsymbol{y}_i^{\text{rep}}, \boldsymbol{a}_i; \boldsymbol{y}_i^{\text{obs}})$ denote the loss for guessing $\boldsymbol{a}_i$ when $\boldsymbol{y}_i^{\text{rep}}$ is obtained and $\boldsymbol{y}_i^{\text{obs}}$ is observed. The criterion is defined as follows:

$$
\begin{aligned}
L(\boldsymbol{Y}^{\text{rep}}, \boldsymbol{a};\ \boldsymbol{Y}^{\text{obs}}) &= \sum_{i=1}^{n} E_{\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{Y}^{\text{obs}}} L(\boldsymbol{y}_i^{\text{rep}}, \boldsymbol{a}_i;\ \boldsymbol{y}_i^{\text{obs}}) \\
&= \sum_{i=1}^{n} \left[E_{\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{Y}^{\text{obs}}} L(\boldsymbol{y}_i^{\text{rep}}, \boldsymbol{a}_i) + k L(\boldsymbol{y}_i^{\text{obs}}, \boldsymbol{a}_i) \right] \qquad (1.1)
\end{aligned}
$$

where $L(\cdot,\cdot)$ denotes a certain loss, and different loss will give different criterion for model choice. $E_{\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{Y}^{\text{obs}}} L(\cdot,\cdot)$ indicates the conditional expectation of $L(\cdot,\cdot)$ which is taken with respect to the conditional distribution $p(\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{Y}^{\text{obs}})$. This equation rewards closeness to $\boldsymbol{y}_i^{\text{rep}}$ but also

to $\boldsymbol{y}_i^{\text{obs}}$, $\boldsymbol{a}_i$ is viewed as a compromise action. The domain $\mathcal{A}$ for $\boldsymbol{a}_i$ needs not to concur with the support of $\boldsymbol{y}_i$. For instance, if $\boldsymbol{y}_i$ is a p-dimensional vector of discrete data, say p Poisson variables, $\mathcal{A}$ would be $\underbrace{R^+ \otimes \cdots \otimes R^+}_{p}$. When the mean of $\boldsymbol{y}_i$ exists, $\mathcal{A}$ will typically be the space of the mean. The scalar k in equation (1.1) indicates the relative regret for departure from $\boldsymbol{Y}^{\text{obs}}$ as compared with departure from $\boldsymbol{Y}^{\text{rep}}$. When $k=0$, this criterion can be viewed as finding a better guess $\boldsymbol{a}$ for $\boldsymbol{Y}^{\text{rep}}$. By using the expected squared Euclidean distance (Ibrahim and Laud, 1994) as the loss function $L(\cdot,\cdot)$, the criterion can be defined as

$$\begin{aligned} L(\boldsymbol{Y}^{\text{rep}},\boldsymbol{a};\boldsymbol{Y}^{\text{obs}}) &= \sum_{i=1}^{n} E_{\boldsymbol{y}_i^{\text{rep}}|\boldsymbol{Y}^{\text{obs}}} L(\boldsymbol{y}_i^{\text{rep}},\boldsymbol{a}_i;\boldsymbol{y}_i^{\text{obs}}) \\ &= \sum_{i=1}^{n} E_{\boldsymbol{y}_i^{\text{rep}}|\boldsymbol{Y}^{\text{obs}}} (\boldsymbol{y}_i^{\text{rep}}-\boldsymbol{a}_i)^{\mathrm{T}}(\boldsymbol{y}_i^{\text{rep}}-\boldsymbol{a}_i) + \\ &\quad \sum_{i=1}^{n} k(\boldsymbol{y}_i^{\text{obs}}-\boldsymbol{a}_i)^{\mathrm{T}}(\boldsymbol{y}_i^{\text{obs}}-\boldsymbol{a}_i) \end{aligned} \tag{1.2}$$

The minimizing $\boldsymbol{a}_i$ is $(1+k)^{-1}(\boldsymbol{\mu}_i + k\boldsymbol{y}_i^{\text{obs}})$, where $\boldsymbol{\mu}_i = E(\boldsymbol{y}_i^{\text{rep}} \mid \boldsymbol{Y}^{\text{obs}})$. Insert these $\boldsymbol{a}_i$ into equation (1.2), and let $v = k/(k+1)$, we get the L_v measure (Ibrahim et al, 2001), which is given by

$$\begin{aligned} L_v(\boldsymbol{Y}^{\text{obs}}) &= \sum_{i=1}^{n}\left[E_{\boldsymbol{y}_i^{\text{rep}}|\boldsymbol{Y}^{\text{obs}}} (\boldsymbol{y}_i^{\text{rep}}-\boldsymbol{\mu}_i)^{\mathrm{T}}(\boldsymbol{y}_i^{\text{rep}}-\boldsymbol{\mu}_i) + v(\boldsymbol{\mu}_i-\boldsymbol{y}_i^{\text{obs}})^{\mathrm{T}}(\boldsymbol{\mu}_i-\boldsymbol{y}_i^{\text{obs}})\right] \\ &= \sum_{i=1}^{n}\left\{ \operatorname{tr}(\operatorname{Var}(\boldsymbol{y}_i^{\text{rep}} \mid \boldsymbol{Y}^{\text{obs}})) + v(\boldsymbol{\mu}_i-\boldsymbol{y}_i^{\text{obs}})^{\mathrm{T}}(\boldsymbol{\mu}_i-\boldsymbol{y}_i^{\text{obs}})\right\} \end{aligned} \tag{1.3}$$

From the definition, the L_v measure can be viewed as a combination of two terms. The first one is the predicted variance which can be viewed as a penalty, and the second one is the predicted bias which can be viewed

as a goodness-of-fit measure. Therefore, the model with the smallest value of the L_v measure will be selected.

1.3 Outline of the Book

In this book, we focus on model selection for several different kinds of SEMs. To the best of our knowledge, Bayes factor and DIC are the most popular methods for model selection of SEMs. Due to the reasons given in the section of overview, they have some limitations. Hence, there is a need to develop an efficient and simple approach to deal with the problem of model selection in structural equation modeling. In Chapter 2, the L_v measure is applied to nonlinear SEMs. In addition, the calibration distribution of the L_v measure is discussed. In Chapter 3, the L_v measure is further applied to nonlinear SEMs with mixed continuous and ordinal categorical responses. In Chapter 4, considering the existence of hierarchical observations in real applications, the L_v measure, together with the calibration distribution, is used for model selection of two-level SEMs. In Chapter 5, a finite mixture of SEMs with unknown number of components is considered for the analysis of heterogeneous data. The L_v measure is used to perform the model selection of mixture SEMs. Simulation studies and real data analyses are conducted to demonstrate the proposed methodologies in these chapters. Besides, to address the performances of different model selection methods, Bayes factor and DIC are also computed for model selection in this book. Conclusions and further developments are presented in Chapter 6, and technical details are given in the Appendix.

Chapter 2 L_v Measure for Nonlinear Structural Equation Models

2.1 Introduction

Model selection is an important issue in data analysis. Recently, many methods for model assessment and model selection have been developed. However, for structural equation models (SEMs), it is a difficult problem due to the complexity of SEMs. To deal with the problem, Bayes factor was proposed for model selection in structural equation modeling (see Jedidi et al., 1997; Lee and Song, 2001, 2003b; Lee, 2007). But as pointed out by Ibrahim et al (2001) and Kass and Raftery (1995), this method relies on posterior model probabilities, and proper prior distributions of unknown parameters are needed. Therefore, it is usually a major task to specify the prior distributions for all models under consideration, in particular when the model space is large. Moreover, Bayes factor is generally sensitive to the choice of prior distributions, and its computational burden is heavy. Alternatively, criterion-based methods are attractive in the sense that they do not require proper prior distributions in general, and the computational burden is much light compared with Bayes factor. There are many criterion-based methods for model selection, such as Akaike information criterion (AIC) (Akaike, 1973, 1981), Bayesian information criterion (BIC) (Schwarz, 1978), and deviance information criterion (DIC) (Spiegelhalter et al., 2002). AIC and BIC are statistics for model assessment and selection based on maximum likelihood estimates. In this chapter, a statistic called the L_v measure (see Gelfand and Ghosh, 1998; Ibrahim et al., 2001; Chen et al.,

2004) will be applied to model selection in nonlinear SEMs. As discussed in Ibrahim et al (2001), the L_v measure can be written as a sum of two components, one is related to the reliability of the prediction, and the other is related to the discrepancy between the prediction and the observed data. It can be viewed as a Bayesian goodness-of-fit statistic, and can be used as a criterion for model assessment and selection. By using MCMC method, the computation of the L_v measure is quite easy after obtaining the estimates of unknown parameters and latent variables. To compare the performance of different model selection methods, Bayes factor and DIC will also be computed for model selection in this chapter.

The remainder of this chapter is divided into six sections. In Section 2.2, a brief review of the L_v measure for model selection will be given. In Section 2.3, a nonlinear SEM will be specified. In Section 2.4, the L_v measure for model selection of nonlinear SEMs will be introduced. In Section 2.5, a simulation study is presented to demonstrate the performance of the L_v measure. In Section 2.6, a real example is analyzed to illustrate the methodology. A discussion is given in Section 2.7.

2.2 Brief Review of the L_v Measure

Let $\boldsymbol{Y}^{\text{obs}} = (\boldsymbol{y}_1^{\text{obs}}, \cdots, \boldsymbol{y}_n^{\text{obs}})$ be a matrix of observations, and $\boldsymbol{Y}^{\text{rep}} = (\boldsymbol{y}_1^{\text{rep}}, \cdots, \boldsymbol{y}_n^{\text{rep}})$, which has the same distribution with $\boldsymbol{Y}^{\text{obs}}$, be the future value of an imagined replicate experiment. Suppose that for the observations in $\boldsymbol{Y}^{\text{obs}}$, a class of models denoted by $\{M_t, t = 0,1,\cdots,T\}$ are considered. Under a certain model M_t, let $\boldsymbol{\theta}$ be the parameter vector that contains all unknown parameters in the model, and $\boldsymbol{a} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ be an estimate trying to accommodate both $\boldsymbol{Y}^{\text{obs}}$ and $\boldsymbol{Y}^{\text{rep}}$. Then a minimum posterior predictive loss for this model was proposed (see Gelfand and Ghosh, 1998):

$$L_k(\boldsymbol{Y}^{\text{obs}}, M_t) = \sum_{i=1}^{n} \min_{\boldsymbol{a}_i} E_{\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{Y}^{\text{obs}}, M_t} L(\boldsymbol{y}_i^{\text{rep}}, \boldsymbol{a}_i; \boldsymbol{y}_i^{\text{obs}})$$

$$= \sum_{i=1}^{n} \min_{\boldsymbol{a}_i} \left\{ E_{\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{Y}^{\text{obs}}, M_t} L(\boldsymbol{y}_i^{\text{rep}}, \boldsymbol{a}_i) + kL(\boldsymbol{y}_i^{\text{obs}}, \boldsymbol{a}_i) \right\} \quad (2.1)$$

where k is a weight that indicates the trade-off between the departure from $\boldsymbol{y}_i^{\text{obs}}$ and the departure from $\boldsymbol{y}_i^{\text{rep}}$, $L(\cdot,\cdot)$ denotes a certain loss, and different loss will give different criterion for model choice. $L(\boldsymbol{y}_i^{\text{obs}}, \boldsymbol{a}_i; \boldsymbol{y}_i^{\text{obs}})$ can be interpreted as the loss for guessing $\boldsymbol{a}_i$ when $\boldsymbol{y}_i^{\text{rep}}$ is obtained and $\boldsymbol{y}_i^{\text{obs}}$ is observed. In equation (2.1), by using the Euclidean distance defined in Ibrahim and Laud (1994), the L_v measure (see Ibrahim et al, 2001) for model M_t is defined as follows:

$$L_v(\boldsymbol{Y}^{\text{obs}}, M_t) = \sum_{i=1}^{n} \text{tr}(\text{Var}(\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{Y}^{\text{obs}}, M_t)) +$$

$$v \sum_{i=1}^{n} (\boldsymbol{\mu}_i - \boldsymbol{y}_i^{\text{obs}})^{\text{T}} (\boldsymbol{\mu}_i - \boldsymbol{y}_i^{\text{obs}}) \quad (2.2)$$

where $\boldsymbol{\mu}_i = E(\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{Y}^{\text{obs}}, M_t)$, and $v = \frac{k}{k+1}, 0 \leqslant v < 1$. From equation (2.1), $k \in [0, \infty)$ is a trade-off between two losses. $k = 1$ means equal weights, which makes $v = 0.5$. Therefore, in this chapter, we will consider the L_v measure with v equals to 0.5. The conditional variance and expectation in equation (2.2) are taken with respect to the posterior predictive distribution $(\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{Y}^{\text{obs}}, M_t)$, which is defined by

$$p(\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{Y}^{\text{obs}}, M_t) = \int p(\boldsymbol{y}_i^{\text{rep}} | \boldsymbol{\theta}, M_t) p(\boldsymbol{\theta} | \boldsymbol{Y}^{\text{obs}}, M_t) \text{d}\boldsymbol{\theta}$$

From its definition, the L_v measure can be viewed as a Bayesian goodness-of-fit statistic, which measures the performance of a model by a combination of how close its predictions are to the observed data and the

variability of the predictions. The model with the smallest value of the L_v measure will be selected. Specifically, let $P=\sum_{i=1}^{n}\mathrm{tr}(\mathrm{Var}(\boldsymbol{y}_i^{\mathrm{rep}}\mid\boldsymbol{Y}^{\mathrm{obs}},M_t))$ and $G=\sum_{i=1}^{n}(\boldsymbol{\mu}_i-\boldsymbol{y}_i^{\mathrm{obs}})^{\mathrm{T}}(\boldsymbol{\mu}_i-\boldsymbol{y}_i^{\mathrm{obs}})$, then $L_v(\boldsymbol{Y}^{\mathrm{obs}},M_t)=P+v\times G$, where P can be viewed as a penalty term, and G is an error sum of squares and can be viewed as a goodness-of-fit measure. For over-fitted model, P will decrease while G will increase; for underestimated model, P will increase while G will decrease. Therefore, complexity is penalized and a parsimonious choice is encouraged.

2.3 Model Description

Let $\boldsymbol{y}_i$, for $i=1,\cdots,n$, be a $p\times1$ random vector of observed variables, and $\boldsymbol{Y}=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n)$. The nonlinear SEM denoted by M is defined by

$$M:\quad \boldsymbol{y}_i=\boldsymbol{u}+\boldsymbol{\Lambda}\boldsymbol{\omega}_i+\boldsymbol{\varepsilon}_i \tag{2.3}$$

$$\boldsymbol{\eta}_i=\boldsymbol{\Pi}\boldsymbol{\eta}_i+\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i)+\boldsymbol{\delta}_i \tag{2.4}$$

where $\boldsymbol{u}$ is a $p\times1$ mean vector; $\boldsymbol{\omega}_i$ is a $q\times1$ vector of latent variables; $\boldsymbol{\varepsilon}_i$ is a $p\times1$ random vector of error terms, and is independent of $\boldsymbol{\omega}_i$; $\boldsymbol{\omega}_i=(\boldsymbol{\eta}_i^{\mathrm{T}},\boldsymbol{\xi}_i^{\mathrm{T}})^{\mathrm{T}}$ is partition of $\boldsymbol{\omega}_i$ into endogenous and exogenous latent vectors $\boldsymbol{\eta}_i(q_1\times1)$ and $\boldsymbol{\xi}_i(q_2\times1)$, respectively; $\boldsymbol{\Pi}$ and $\boldsymbol{\Gamma}$ are matrices of unknown regression coefficients; $\boldsymbol{F}(\cdot)=(f_1(\cdot),\cdots,f_r(\cdot))^{\mathrm{T}}$ is a vector-valued function with differentiable functions $f_1(\cdot),\cdots,f_r(\cdot)$, and $r\geqslant q_2$; $\boldsymbol{\delta}_i$ is a $q_1\times1$ random vector of error terms, and is independent of $\boldsymbol{\xi}_i$. We assume that, for $i=1,\cdots,n$

$$\boldsymbol{\varepsilon}_i\sim N(\boldsymbol{0},\boldsymbol{\Psi}_{\varepsilon}),\quad \boldsymbol{\xi}_i\sim N(\boldsymbol{0},\boldsymbol{\Phi}),\quad \boldsymbol{\delta}_i\sim N(\boldsymbol{0},\boldsymbol{\Psi}_{\delta}) \tag{2.5}$$

where $\boldsymbol{\Psi}_{\varepsilon}=\mathrm{diag}(\psi_{\varepsilon1},\cdots,\psi_{\varepsilon p})$ and $\boldsymbol{\Psi}_{\delta}=\mathrm{diag}(\psi_{\delta1},\cdots,\psi_{\delta q_1})$ are diagonal matrices.

Let $\boldsymbol{G}(\boldsymbol{\omega}_i)=(\boldsymbol{\eta}_i^{\mathrm{T}},\boldsymbol{F}(\boldsymbol{\xi}_i)^{\mathrm{T}})^{\mathrm{T}}$, and $\boldsymbol{\Lambda}_\omega=(\boldsymbol{\Pi},\boldsymbol{\Gamma})$,then equation (2.4) can be rewritten as $\boldsymbol{\eta}_i=\boldsymbol{\Lambda}_\omega\boldsymbol{G}(\boldsymbol{\omega}_i)+\boldsymbol{\delta}_i$. Moreover, Let $\boldsymbol{\Lambda}_\eta$ and $\boldsymbol{\Lambda}_\xi$ be the submatrices of $\boldsymbol{\Lambda}$ corresponding to $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$, respectively. And let $\boldsymbol{\Pi}_0=\boldsymbol{I}-\boldsymbol{\Pi}$, which is assumed to be nonsingular, then model M can be written as

$$\boldsymbol{y}_i=\boldsymbol{u}+\boldsymbol{\Lambda}\boldsymbol{\Pi}_0^{-1}(\boldsymbol{\Gamma}\boldsymbol{F}(\boldsymbol{\xi}_i)+\boldsymbol{\delta}_i)+\boldsymbol{\Lambda}\boldsymbol{\xi}_i+\boldsymbol{\varepsilon}_i \tag{2.6}$$

The Bayesian approach (see Dunson, 2000; Lee and Song, 2004; Lee, 2007) can be applied to the estimation of this nonlinear SEM. Here, we will focus on model comparison based on the L_v measure. For convenience, the following notations are used. Let $\boldsymbol{Y}^{\mathrm{obs}}=(\boldsymbol{y}_1^{\mathrm{obs}},\cdots,\boldsymbol{y}_n^{\mathrm{obs}})$ be the observed continuous data, where $\boldsymbol{y}_i^{\mathrm{obs}}=(y_{i1}^{\mathrm{obs}},\cdots,y_{ip}^{\mathrm{obs}})^{\mathrm{T}}(i=1,\cdots,n)$ is the ith column of $\boldsymbol{Y}^{\mathrm{obs}}$, and $\boldsymbol{Y}^{\mathrm{rep}}=(\boldsymbol{y}_1^{\mathrm{rep}},\cdots,\boldsymbol{y}_n^{\mathrm{rep}})$ be the replicated data set which has the same distribution with $\boldsymbol{Y}^{\mathrm{obs}}$, where $\boldsymbol{y}_i^{\mathrm{rep}}=(y_{i1}^{\mathrm{rep}},\cdots,y_{ip}^{\mathrm{rep}})^{\mathrm{T}}$. Let $\boldsymbol{\Omega}=(\boldsymbol{\omega}_1,\cdots,\boldsymbol{\omega}_n)$ be the matrix of latent variables. Moreover, let $\boldsymbol{\Omega}_1=(\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_n)$ and $\boldsymbol{\Omega}_2=(\boldsymbol{\xi}_1,\cdots,\boldsymbol{\xi}_n)$ be the submatrices of $\boldsymbol{\Omega}$ corresponding to $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$, respectively. Furthermore, let $\boldsymbol{G}=(\boldsymbol{G}(\boldsymbol{\omega}_1),\cdots,\boldsymbol{G}(\boldsymbol{\omega}_n))$, and let $\boldsymbol{\theta}$ be the vector that contains all the unknown elements in $\boldsymbol{u}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}_\varepsilon$, $\boldsymbol{\Pi}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}_\delta$ in the model defined by equations (2.3) and (2.4). Finally, let Θ be the space of the parameter vector $\boldsymbol{\theta}$, and Ξ be the space of the latent variables $\boldsymbol{\xi}_i$, for $i=1,\cdots,n$.

2.4 L_v Measure for Nonlinear Structural Equation Models

2.4.1 Definition of the L_v measure

In this part, the L_v measure will be applied to the nonlinear SEM, M,